

O Uso da Inteligência Artificial para Auxiliar na Predição de Abandono de Alunos em uma Escola de Ensino Regular

Bianca S. da Silva,¹ Joyce de L. C. Rangel²

DCC/UFRRJ, Nova Iguaçu, RJ

Vanessa A. Silva,³ Marcelo D. Cruz⁴

PPGMMC/UFRRJ, Seropédica, RJ

Ronaldo M. Gregório⁵

PPGMMC/UFRRJ, Seropédica, RJ

Resumo A Evasão Escolar é um desafio que impacta sistemas educacionais em todo o mundo. Este trabalho utiliza algoritmos de Inteligência Artificial e Aprendizado de Máquina para enfrentar o problema, aplicando técnicas capazes de analisar dados históricos e antecipar tendências futuras. Especificamente, utilizamos Redes Neurais Artificiais para identificar padrões e compreender os fatores que levam os alunos a abandonar os estudos. A coleta de dados contou com a colaboração de alunos da Educação de Jovens e Adultos do Novo Ensino Médio (EJANEM), que, em algum momento de suas trajetórias escolares, precisaram interromper os estudos. Esses dados foram obtidos por meio de formulários preenchidos pelos alunos e comparados com dados coletados de alunos do Ensino Regular.

Palavras-chave. Evasão Escolar. Inteligência Artificial. Aprendizado de Máquina. Modelo de Classificação. Aprendizado Supervisionado. Redes Neurais.

1 Introdução

A Evasão Escolar é um desafio que afeta sistemas educacionais em todo o mundo. Diversas causas contribuem para esse fenômeno, incluindo problemas familiares, necessidade de contribuir financeiramente para as despesas de casa, gravidez na adolescência e baixa perspectiva, entre outros [4][3]. Compreender e prever os fatores que levam os alunos a abandonar os estudos é essencial para desenvolver estratégias eficazes de prevenção e intervenção.

Para a coleta de dados, contamos com a colaboração de uma escola da Rede Estadual de Ensino do Rio de Janeiro, onde foi realizado um levantamento sobre os alunos desistentes ao longo de dez anos. Contudo, logo percebemos que a escola não dispunha de informações relevantes sobre esses alunos. Diante desse desafio, foi necessária uma mudança de estratégia: contamos com a participação dos alunos da Educação de Jovens e Adultos do Novo Ensino Médio (EJANEM), uma modalidade de ensino voltada ao resgate acadêmico de jovens e adultos. Essa modalidade é oferecida no turno noturno e distribui todos os conteúdos do Ensino Médio ao longo de quatro módulos, totalizando um período de dois anos. O EJANEM chamou nossa atenção por ser representado por um grupo de estudantes com trajetórias educacionais não convencionais. Muitos desses alunos, em algum momento de suas vidas, interromperam os estudos no Ensino Regular. Essa valiosa fonte

¹bia20santanadasilva@gmail.com

²joylcrangel@ufrj.br

³vanessa.mec.ime@gmail.com

⁴madibcruz@gmail.com

⁵doug8780@yahoo.com.br

de dados nos proporcionou a oportunidade de explorar as razões subjacentes à Evasão Escolar e identificar padrões que contribuem para esse fenômeno.

A Inteligência Artificial (IA) é um campo da ciência que visa realizar tarefas de forma semelhante à dos seres humanos [8]. Uma das áreas da IA é o Aprendizado de Máquina, que permite a aprendizagem a partir de dados [8][7]. Esse método baseia-se na ideia de que os algoritmos aprendem com dados anteriores para realizar uma previsão do futuro. Existem vários métodos de Aprendizado de Máquina, como Redes Neurais Artificiais, Árvores de Decisão e Máquinas de Vetores de Suporte (SVM - Support Vector Machine), entre outros [5]. As Redes Neurais Artificiais, em particular, são modelos inspirados no funcionamento dos neurônios do cérebro humano, compostos por camadas de unidades interconectadas por neurônios artificiais [1].

As Redes Neurais Artificiais podem utilizar dados dos alunos, incluindo fatores socioeconômicos, características familiares e outras variáveis relevantes. Ao explorar essa riqueza de dados, o objetivo é desenvolver um modelo capaz de classificar os alunos com base em suas características e prever se estão em risco de evasão ou não. Com essa abordagem, pretendemos fornecer informações significativas para educadores, formuladores de políticas públicas e profissionais da área da educação. A análise dos resultados obtidos pelo modelo de previsão de evasão contribuirá para um entendimento mais profundo das circunstâncias que levam ao problema, permitindo a implementação de medidas preventivas mais direcionadas.

Nas seções subsequentes, descreveremos em detalhes a metodologia utilizada para a coleta e preparação dos dados, a implementação do algoritmo de previsão e a análise dos resultados obtidos. Acreditamos que a aplicação do Aprendizado de Máquina Supervisionado a este contexto específico pode oferecer contribuições valiosas para a compreensão das causas da Evasão Escolar e, conseqüentemente, para a promoção de um ambiente educacional mais inclusivo e bem-sucedido.

2 Motivação

Esta pesquisa busca contribuir para o aprimoramento da gestão educacional, promovendo a permanência dos alunos e, conseqüentemente, seu progresso dentro da instituição escolar. Ao identificar precocemente os alunos em risco de evasão, é possível desenvolver estratégias proativas e individualizadas para evitar a interrupção de seu percurso acadêmico, fortalecendo, assim, o sistema educacional como um todo.

3 Objetivo

O objetivo deste estudo é aplicar Redes Neurais Artificiais a dados de alunos para identificar possíveis casos de abandono escolar. Espera-se que essa abordagem preventiva contribua para a criação de um ambiente escolar mais inclusivo, oferecendo suporte aos gestores educacionais e ajudando a reduzir a evasão.

4 Metodologia

Para obter dados relevantes para a pesquisa, foi proposto um levantamento dos alunos que evadiram ao longo de um período de dez anos em uma escola da Rede Estadual de Ensino do Rio de Janeiro. No entanto, devido à falta de informações disponíveis sobre esses alunos, foi necessária a adoção de uma estratégia alternativa: buscar informações com os alunos do programa de Educação de Jovens e Adultos do Novo Ensino Médio (EJANEM), pois muitos deles eram anteriormente alunos do Ensino Regular e tiveram que interromper seus estudos por algum motivo.

Ao coletar informações dos alunos dessa modalidade de ensino, foi possível obter uma visão mais completa e aprofundada das causas da Evasão Escolar na instituição. Compreender as circunstâncias específicas que levaram à interrupção dos estudos é fundamental para desenvolver estratégias e políticas eficazes de prevenção da Evasão Escolar, além de fornecer suporte adequado aos alunos em situação de risco [6]. Essa abordagem contribui não apenas para a compreensão do problema, mas também permite que os alunos do Ejanem sejam ouvidos e tenham a oportunidade de compartilhar suas experiências e perspectivas. Esse envolvimento promove um ambiente de participação e valorização, no qual suas vozes são consideradas na busca por soluções.

Para realizar o levantamento dos dados, foi proposto o uso de um formulário contendo informações relevantes para o projeto. Esse formulário foi elaborado com base em estudos da literatura [4][3]. Inicialmente, a pesquisa seria realizada online, com um formulário enviado por e-mail, permitindo que os alunos respondessem no momento e local mais conveniente. Contudo, surgiram os primeiros contratempos: muitos alunos enfrentaram dificuldades para compreender as perguntas ou não demonstraram interesse em respondê-las. Essa abordagem online resultou em uma adesão limitada, não atingindo o objetivo esperado de obter um número significativo de respostas completas e precisas.

Diante desses obstáculos, foi necessária uma abordagem alternativa mais eficaz. Um formulário impresso foi elaborado e distribuído em todas as turmas dos quatro módulos do Ejanem. O método presencial possibilitou alcançar um número maior de alunos, cobrindo aproximadamente oito turmas, com um total de 200 alunos. Os formulários foram preenchidos em sala de aula, permitindo que dúvidas fossem sanadas imediatamente — vale ressaltar que todos os questionários preenchidos foram anônimos. A mudança do formato online para o impresso mostrou-se essencial para aumentar o engajamento dos alunos e permitir uma coleta de dados mais eficiente. Com os formulários impressos em mãos, o próximo passo foi cadastrar as respostas em sistemas ou bancos de dados adequados, garantindo que os dados estivessem organizados e prontos para análise [2].

Concluída essa etapa da pesquisa, tornou-se necessário conhecer os perfis dos alunos que nunca abandonaram a escola, ou seja, alunos do Ensino Regular. Para isso, foi utilizado um conjunto de informações previamente disponível na escola sobre esses alunos. Apesar de muito valiosas, as informações coletadas nas duas modalidades (Ejanem e Ensino Regular) não estavam totalmente alinhadas.

Assim, foi elaborado um novo formulário com perguntas idênticas para alunos do Ensino Regular e do Ejanem, permitindo uma comparação direta entre os dados. Dessa vez, a pesquisa realizada em ambas as modalidades contou com a participação de 176 alunos do Ejanem e 201 alunos do Ensino Regular.

5 Resultados

Inicialmente, a amostra foi composta pela participação de alunos da modalidade Ejanem. Foram utilizadas informações já existentes na escola para comparar os dados dos alunos do Ensino Regular. Por esse motivo, os atributos não estavam totalmente equivalentes, sendo necessário utilizar apenas algumas colunas do dataset. Essa primeira abordagem foi fundamental para testar o modelo, o que nos garantiu uma acurácia de 79%, mesmo com a redução no número de atributos. Após essa etapa inicial, foi utilizado o mesmo formulário para todos os alunos. Foram feitos alguns ajustes, como a transformação de dados categóricos em variáveis numéricas. O atributo "série", em particular, refere-se à série em que o aluno está matriculado atualmente no Ensino Regular e à série abandonada pelos alunos do Ejanem. Abaixo são apresentadas essas transformações:

- **sexo:**

Masculino = 1;

Feminino = 2;

- **idade:**

Número inteiro correspondente.

- **renda_mensal:**

inferior a um salário mínimo = 1;

um salário mínimo = 2;

entre um e dois salários mínimos = 3;

entre dois e três salários mínimos = 4;

entre três e quatro salários mínimos = 5;

acima de quatro salários mínimos = 6;

- **num_moradores (incluindo você):**

Número inteiro correspondente.

- **turno:**

Manhã = 1;

Tarde = 2;

Noite = 3.

- **serie**

Fundamental;

6° ano = 1;

7° ano = 2;

8° ano = 3;

9° ano = 4;

Ensino Médio;

1° ano = 5;

2° ano = 6;

3° ano = 7;

antes do 6° ano = 8.

- **responsavel_por_alguem:**

sim = 1;

não = 2;

- **nucleo_familiar_afeta:**

sim = 1;

não = 2;

da capacidade de generalização do modelo em dados não observados.

Nesse caso, foram divididos em 3 folders e a média obtida foi de 0.869. Isso significa que o modelo acerta o abandono e não abandono em 87% das vezes (acurácia de 87%). Dos casos em que o modelo classificou como abandono, ele acertou em 86% das vezes (precisão de 86%), e de todos os casos de abandono contidos nos dados, o modelo classificou corretamente como abandono em 85% das vezes (Recall de 85%). Em suma, os resultados foram bastante significativos.

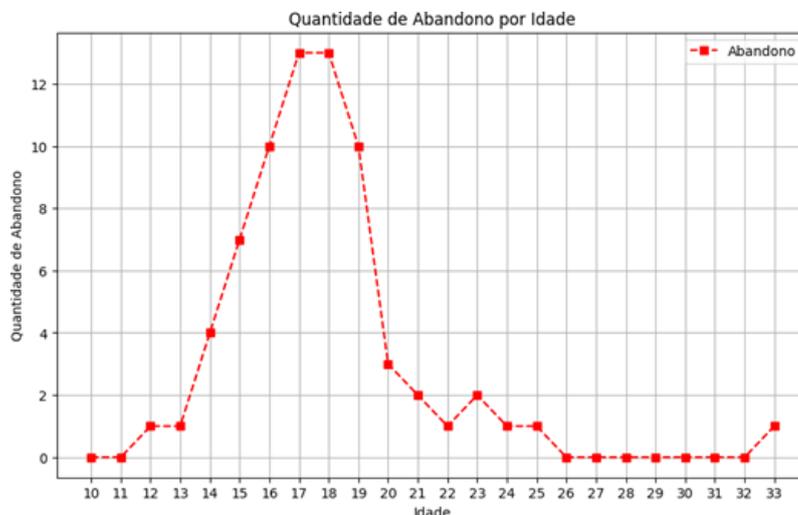


Figura 2: Gráfico de abandono por idade. Fonte: dos autores.

6 Considerações Finais

Foi possível observar uma melhora significativa do modelo após a reestruturação do formulário. É importante considerar múltiplos fatores ao prever a Evasão Escolar, uma vez que essas variáveis aparentemente distintas podem estar interligadas e influenciar conjuntamente o resultado final. Essa descoberta ressalta a complexidade do problema da Evasão Escolar e a necessidade de abordagens analíticas avançadas. O modelo de Classificação demonstrou uma excelente acurácia de 87% na previsão da Evasão Escolar. Isso significa que o modelo aprendeu bem mediante as informações fornecidas e que é capaz de acertar a predição em 87% dos casos analisados.

Isso reforça a eficácia dos métodos de Aprendizado de Máquina em fornecer informações importantes para auxiliar instituições educacionais na tomada de decisões. No entanto, ressaltamos a necessidade contínua de refinamento e melhoria do modelo, bem como a consideração de outros indicadores e variáveis relevantes que podem aprimorar ainda mais sua capacidade preditiva.

Agradecimentos

À Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro – RJ/Brasil (FAPERJ), pelo apoio financeiro ao projeto de pesquisa E-26/210.191/2022, do qual o presente estudo faz parte.

Referências

[1] C. M. Bishop. **Neural Networks for Pattern Recognition**. Oxford University Press, 1995.

- [2] M. M. C. de Carvalho. “Evasão Escolar no Brasil: Uma Perspectiva Multidimensional”. Em: **Biblioteca Virtual da FAFESP** (2014).
- [3] R. M. Dutra. “O uso de inteligência artificial para predição de evasão na rede Doctum de ensino”. Dissertação de mestrado. Instituto Doctum de Educação e Tecnologia, Faculdades Integradas de Caratinga - FIC, 2015.
- [4] L. S. Mendes e F. C. Souza. “Análise das causas da evasão escolar no ensino fundamental”. Em: **Educação em Foco** 25.2 (2021), pp. 89–102. DOI: 10.5752/P.2317-788X.2021v25n2p89.
- [5] S. Haykin. **Neural Networks and Learning Machines**. Prentice Hall, 2008.
- [6] H. S. Carvalho e M. I. S. Castanho. “A evasão escolar no ensino médio: Análise de uma realidade”. Em: **Revista Póescrito** 13 (2022). DOI: 10.15448/2179-8435.2022.1.40630. URL: <https://revistaseletronicas.pucrs.br/index.php/poescrito/article/view/40630>.
- [7] T. M. Mitchell. **Machine Learning**. McGraw-Hill, 1997. ISBN: 0070428077.
- [8] S. Russell e P. Norvig. **Inteligência Artificial**. 3. ed. Elsevier, 2013.
- [9] J. VanderPlas. **Python Data Science Handbook: Essential Tools for Working with Data**. O’Reilly Media, 2016.